

Il deposito, la valorizzazione e la conservazione delle tesi di dottorato nell'esperienza di Magazzini digitali: un contributo per la ricerca e l'accesso

Chiara Storti^(a)

a) Biblioteca Nazionale Centrale di Firenze

Contact: Chiara Storti, storti.chiara@gmail.com

Received: 19 October 2018; Accepted: 28 October 2018; First Published: 15 January 2019

ABSTRACT

This contribution traces the history of the legal deposit of Doctoral thesis in Italy from the National Central Library of Florence's point of view, in order to constitute a guide for research and access. Since 2006 the National Library of Florence manages the national digital preservation and long-term access service called "Magazzini Digitali" ("Digital Stocks").

ACKNOWLEDGEMENTS

Essendo l'ultima, in ordine di tempo, ad occuparmi di Magazzini Digitali, è doveroso sottolineare che il contributo che segue presenta un lavoro svolto in gran parte dalle persone che mi hanno preceduto e che, a vario titolo, continuano con passione a collaborare al progetto. Tra questi, certamente, Giovanni Bergamin e Maurizio Messina. Sono anche molto grata ai miei attuali colleghi, sia della BNCF che non, che insieme a me portano avanti queste attività cercando di garantire un servizio sempre migliore: Cristiano Corsani, Maria Grazia Pepe, Massimiliano Randazzo, Raffaele Messuti, Emanuele Bellini. Ad Andrea Marchitelli, Raffaele Messuti e Giovanni Bergamin devo anche i ringraziamenti per l'attenta rilettura del testo e i preziosi consigli.

KEYWORDS

Doctoral thesis; Legal Deposit; Long Term Preservation; Digital Preservation; Digital Library.

CITATION

Storti, C. "Il deposito, la valorizzazione e la conservazione delle tesi di dottorato nell'esperienza di Magazzini digitali: un contributo per la ricerca e l'accesso." *J LIS.it* 10, 1 (January 2019): 114–124. DOI: [10.4403/jlis.it-12526](https://doi.org/10.4403/jlis.it-12526).

Storia e riferimenti normativi

Questo breve contributo sulla gestione delle tesi di dottorato in Italia, che si soffermerà in particolare sulle tesi in formato digitale, non può non aprirsi che ripercorrendo alcuni punti salienti della storia del Dottorato di ricerca nel nostro Paese e della normativa sul deposito delle tesi.

In questa introduzione non vi è alcuna pretesa di esaustività, ma è necessario conoscere il contesto di produzione per meglio comprendere questi oggetti culturali che, come si vedrà, si presentano multiformi oltre che per le tipologie di supporto, anche per la particolare natura documentaria: prodotto della ricerca scientifica, atto valido ai fini del rilascio del titolo di Dottore di Ricerca, documento con dignità di pubblicazione.¹

Il punto di vista è quello della Biblioteca Nazionale Centrale di Firenze (d'ora in poi BNCf), una delle due istituzioni italiane, insieme alla Biblioteca Nazionale Centrale di Roma (d'ora in poi BNCr), depositarie di queste risorse.

Come è noto, il Dottorato di ricerca in Italia viene istituito, con un certo ritardo rispetto agli altri Paesi europei, con il D.P.R. 11 luglio 1980, n. 382 "Riordinamento della docenza universitaria, relativa fascia di formazione nonché sperimentazione organizzativa e didattica" - TITOLO III, Capo II "Dottorato di ricerca"; all'art. 73 si legge:

"Il rilascio del titolo di dottore di ricerca è subordinato al deposito di copie, anche non stampate, dei lavori sulla base dei quali il titolo è stato conseguito presso le Biblioteche nazionali di Roma e Firenze, che ne devono assicurare la pubblica consultabilità per non meno di trenta anni. I testi di cui sopra devono essere corredati dalla relazione dei commissari, incluse le eventuali relazioni di minoranza."

Inizialmente il soggetto obbligato al deposito della tesi di dottorato presso le due Biblioteche Nazionali Centrali era il Dottore di ricerca. L'obbligo passa in carico alle Università con il Decreto Ministeriale 30 aprile 1999 n. 224 "Regolamento in materia di dottorato di ricerca":

(art. 6 c. 11) "Il titolo è rilasciato dal rettore dell'università che, a richiesta dell'interessato, ne certifica il conseguimento. Successivamente al rilascio del titolo, l'università medesima cura il deposito di copia della tesi finale presso le biblioteche nazionali di Roma e Firenze."

Delle tesi finali si occupa anche il più recente Decreto Ministeriale 8 febbraio 2013 n. 45 "Regolamento recante modalità di accreditamento delle sedi e dei corsi di dottorato e criteri per la istituzione dei corsi di dottorato da parte degli enti accreditati":

(art. 14 c. 1) "Il Ministero cura la costituzione e l'aggiornamento di un'anagrafe nazionale dei dottorati di ricerca e di una banca dati delle tesi di dottorato."

(art. 14 c. 3) "Entro trenta giorni dalla discussione e approvazione della tesi l'università o il soggetto promotore depositano copia della tesi, in formato elettronico, nella banca dati ministeriale. Previa autorizzazione del collegio dei docenti, possono essere rese indisponibili parti della tesi in relazione

¹ Come meglio spiegato nel paragrafo "Prospettive", la pubblicità delle tesi di dottorato è determinata ex-lege soltanto in Italia. Altrove il tema della natura pubblica delle tesi è maggiormente dibattuto (Shirazi 2018).

all'utilizzo di dati tutelati da segreto industriale ai sensi della normativa vigente in materia. Resta fermo l'obbligo del deposito della tesi presso le biblioteche nazionali centrali di Roma e di Firenze.”

Una banca dati delle tesi di dottorato a cura del Ministero non è mai stata istituita² e, nonostante l'obbligatorietà del deposito presso le Biblioteche Nazionali Centrali non sia mai venuto a mancare, a causa delle vicende che andremo a descrivere, siamo costretti ad ammettere che in Italia è assai complicato riuscire ad ottenere un elenco completo e univoco delle tesi di dottorato pubblicate dalla metà degli anni ottanta ad oggi.

Le tesi in formato elettronico: dai CD-ROM ai repository istituzionali

Fino al 2006 il deposito delle tesi di dottorato avveniva in formato esclusivamente cartaceo, ma è proprio in quell'anno che inizia l'inversione di tendenza che porterà progressivamente ad abbandonare quasi completamente il deposito delle tesi cartacee.³ Nel 2006 è ormai un dato di fatto che le tesi vengano prodotte nativamente in formato elettronico e infatti il MIUR, con nota del 28 luglio 2006 n. 1420, prevede la possibilità per gli atenei di inviare le tesi alle Biblioteche Nazionali di Roma e Firenze tramite supporto ottico (DVD o CD-ROM). Nell'aprile dello stesso anno, nell'ambito della Commissione biblioteche della CRUI, viene costituito il gruppo Open Access che si occupa di rendere disponibili i prodotti della ricerca scientifica, di cui le tesi di dottorato sono un tassello importante. Il gruppo individua nella costituzione dei repository istituzionali uno dei metodi preferiti per sostenere la pubblicazione in Open Access. D'intesa con il MIBAC e le due Biblioteche Nazionali Centrali viene messa a punto un'apposita procedura di deposito delle tesi interamente telematica: nasce il prototipo di Magazzini Digitali e, a meno di un anno di distanza dalla precedente, il MIUR emette una nuova nota (n. 1746 del 20 luglio 2007) con cui autorizza le Università a inviare le tesi secondo tale procedura telematica, che “assicura sia la conservazione delle tesi per un tempo più lungo sia un modello di catalogazione e classificazione conforme ai più recenti standard internazionali”.

La pratica dell'uso dei CD-ROM viene però abbandonata molto più lentamente addirittura di quello della carta, tanto che ancora nel 2017 sono arrivate in BNCF tra le 2.000 e le 3.000 tesi su questo supporto.⁴

² Al momento dell'emanazione del Decreto Ministeriale 8 febbraio 2013 n. 45 da più parti si mise in discussione l'utilità di creare una banca dati ministeriale delle tesi di dottorato che, sostanzialmente, avrebbe duplicato le funzionalità di Magazzini Digitali. Sulla costituzione della banca dati nazionale delle tesi di dottorato alcune considerazioni interessanti in Galimberti (2013). Si può inoltre citare la piattaforma <http://www.pubblitesi.it/> che, come si legge sul sito, “è un sistema documentale, finalizzato alla raccolta ed alla diffusione elettronica delle migliori Tesi di Laurea e di Dottorato, presentate negli Atenei italiani, realizzato dal CNR e dall'Inforav (Istituto, senza fini di lucro, per lo sviluppo e la gestione avanzata dell'informazione), con il Patrocinio del MIUR, del CNR e di Confindustria Servizi Innovativi e Tecnologici, ed il contributo iniziale dell'AICA (Associazione Italiana per l'Informatica ed il Calcolo Automatico) e della Fondazione Cassa di Risparmio di Roma”. Tuttavia, per il carattere volontaristico dell'adesione all'iniziativa, questo database non può essere considerato attendibile in termini di completezza.

³ Nel 2017 sono pervenute in BNCF circa 600 tesi su supporto cartaceo; nel 2018, da gennaio ad agosto, se ne contano meno di un centinaio (Fonte BNCF). I dati ISTAT più recenti parlano di oltre 11.000 studenti che ogni anno conseguono il titolo di dottore di ricerca: https://www.istat.it/it/files//2015/01/Dottori-di-ricerca_DEF.pdf.

⁴ Fonte BNCF.

È ormai opinione comune che i supporti ottici non siano adatti alla conservazione, non solo per la rapida obsolescenza hardware e software degli strumenti di riproduzione e fruizione ma soprattutto per la degradazione ancora più rapida del supporto stesso (Allegrezza 2015). Anche per questo motivo, oltre che per le questioni legate alla catalogazione, la loro gestione risulta complessa. Inoltre, la conservazione nel lungo periodo delle risorse digitali è una procedura che non può far conto solo su una determinata tecnologia: modelli di gestione e standard devono essere necessariamente “agnostici” rispetto ai supporti e alle infrastrutture tecnologiche.

La persistenza nell'uso dei CD e DVD-ROM come mezzo per il deposito è dovuto a una serie di cause, in parte riconducibili alle istituzioni depositanti, in parte a quelle depositarie. Le Università possono non essere dotate di repository per l'auto-archiviazione delle tesi o, più di frequente, i repository non hanno i requisiti tecnici per essere sottoposti correttamente a procedure di raccolta automatica (*harvesting*), o ancora non sono state messe a punto le dovute procedure amministrative interne. La BNCR e la BNCF, a loro volta, non hanno avuto le forze umane ed economiche per sostenere le Università in questo passaggio, pur cercando di offrire un servizio il più possibile adeguato. La BNCF, in accordo con la BNCR, ha optato per la creazione di un repository “locale”: dal 2006, a partire dai CD e dai DVD che arrivano in biblioteca, i file delle tesi e i relativi allegati vengono caricati su una installazione locale di EPrints e corredati manualmente da metadati minimali in formato Dublin Core. Il repository locale viene poi sottoposto ad *harvesting* OAI-PMH, come qualsiasi altro repository di Università, per garantire la conservazione a lungo termine di dati e metadati. In questo modo si ovvia certamente al problema del degrado dei supporti ottici che, al limite, possono anche essere scartati, ma permangono altri problemi. Innanzi tutto, non sono mai state definite linee guida precise in relazione alle tipologie e ai formati dei file trasmissibili su CD, per cui ci si trova a dover gestire file la cui fruibilità nell'immediato e, soprattutto nel lungo periodo, non è facile; allo stesso modo, non vi sono indicazioni sui metadati. Questi ultimi vengono ricavati dal frontespizio della tesi e dagli altri materiali di corredo ma, se pur si cerchi di garantire una certa uniformità, non si segue uno standard descrittivo preciso e non si effettua controllo di autorità sui dati.

La BNCF ha ipotizzato anche l'apertura all'esterno di questo repository per effettuare l'upload da remoto dei file delle tesi da parte di quelle Università che non sono dotate di repository istituzionali propri.⁵ Questa soluzione, che comporterebbe un ripensamento del flusso documentale e gestionale oltre che un adeguamento del software per consentire, ad esempio, un login multi-utente e multi-livello, è stata per ora accantonata perché si spera che i tempi siano sufficientemente maturi per richiedere alle Università di dotarsi di propri repository o di renderli rispondenti ai requisiti per l'*harvesting*.

⁵ Il network OpenAIRE, che sostiene la cultura aperta e promuove la libera diffusione e il riuso dei prodotti della ricerca scientifica, suggerisce alle istituzioni non dotate di proprio repository di usare zenodo.org, un Orphan Repository senza costi di sottoscrizione che consente la pubblicazione in Open Access. Per approfondimenti: <https://www.openaire.eu/20-for-researchers/491-where-can-i-find-a-suitable-repository>.

Magazzini digitali e le Linee guida della CRUI

Fra il 2006 e il 2007 nasce il prototipo di Magazzini Digitali,⁶ il servizio di conservazione e accesso a lungo termine delle risorse digitali. Il servizio trova un riconoscimento istituzionale a seguito dell'emanazione del D.P.R. 3 maggio 2006, n. 252 "Regolamento recante norme in materia di deposito legale dei documenti di interesse culturale destinati all'uso pubblico" che all'art. 37 "Modalità di deposito e acquisizione dei documenti diffusi tramite rete informatica" recita:

(c. 1) "Le modalità di deposito dei documenti diffusi tramite rete informatica sono definite con successivo regolamento adottato ai sensi dell'articolo 5, comma 1, della legge 15 aprile 2004, n. 106 [...]"

(c. 2) "Il Ministero promuove forme volontarie di sperimentazione del deposito, di cui al comma 1 [...] mediante la stipulazione di appositi accordi con i soggetti obbligati al deposito. Gli accordi definiscono le modalità tecniche del deposito prevedendo, ove possibile, anche forme automatiche di raccolta, secondo le migliori pratiche e conoscenze internazionali del settore."

Questi sono anche gli anni delle grandi campagne di digitalizzazione della Biblioteca Digitale Italiana,⁷ e fin da subito si pensa che Magazzini Digitali possa essere un contenitore non solo per il *born digital* ma anche per le digitalizzazioni. Indipendentemente dalle sue più recenti evoluzioni, per altro ancora in corso, e dalle problematiche non del tutto risolte emerse in relazione alle difficoltà di gestire contemporaneamente digitale nativo e non nativo, è indubbio che il deposito delle tesi di dottorato digitali tramite *harvesting* sia stato e sia ancora una delle attività principali di Magazzini Digitali. Nell'ottobre 2007 la Commissione biblioteche della CRUI pubblica le "Linee guida per il deposito delle tesi di dottorato negli archivi aperti",⁸ nella cui premessa si legge:

"Poiché le tesi di dottorato nascono in formato elettronico, il gruppo Open Access ha verificato con le due Biblioteche Nazionali centrali la possibilità di effettuare la consegna in formato elettronico tramite *harvesting*⁹ o tramite invio sicuro dei files. La recente circolare del MiUR che autorizza la consegna delle tesi alle Biblioteche Nazionali in formato elettronico e per via telematica sancisce, anche da un punto di vista amministrativo, il ricorso a tale procedura. Le Linee guida nascono dalla necessità di individuare uno standard nazionale nella raccolta ed esposizione dei dati".

⁶ Il progetto Magazzini Digitali prende avvio nel 2006 con un accordo tra BNCf, BNCR e Fondazione Rinascimento Digitale, che ne è anche il principale finanziatore. Nel 2010, a seguito dell'uscita del D.P.R. 252/2006, con apposita Lettera d'Intenti viene formalizzata la collaborazione tra BNCf, BNCR, Fondazione Rinascimento Digitale e Biblioteca Nazionale Marciana (d'ora in poi BNM) per la realizzazione di un "Servizio coordinato di conservazione e di accesso nel lungo periodo delle risorse digitali". A luglio 2018, soprattutto in vista della prossima emanazione del regolamento relativo alle modalità di deposito dei documenti diffusi tramite rete informatica previsto dall'art. 37 c. 1 del D.P.R. 252/2006, si è proceduto all'aggiornamento e al rinnovo della lettera d'intenti tra BNCf, BNCR e BNM (la Fondazione Rinascimento Digitale è stata nel frattempo chiusa). Per ulteriori informazioni: Bergamin e Messina (2010).

⁷ <http://www.iccu.sbn.it/opencms/opencms/it/main/InternetCulturale/BDI/>.

⁸ https://www.crui.it/images/biblioteche/linee_guida_deposito_tesi_dottorato.pdf.

⁹ Per approfondimenti: <https://www.crui.it/procedura-di-harvesting-automatico.html>, ma anche Vignocchi, Bergamin e Messuti (2010) e Vignocchi e Lauriola (2011).

Lo stesso Gruppo OA pubblica successivamente altri due importanti strumenti che, fornendo alle Università indicazioni precise per la costituzione e la gestione dei repository istituzionali (che quasi sempre raccolgono non solo le tesi di dottorato ma anche gli altri prodotti editoriali frutto della ricerca scientifica), facilitano anche le operazioni di *harvesting*: si tratta delle “Linee guida per gli archivi istituzionali”¹⁰ del 2009 e delle “Linee guida per la creazione e la gestione di metadati nei repository istituzionali”¹¹ del 2012.

L’*harvesting* delle tesi di dottorato

A partire dal 2010, l’obbligo di deposito delle tesi di dottorato presso le due Biblioteche Nazionali Centrali può essere assolto aderendo al servizio di *harvesting* automatico da repository istituzionale gestito dalla BNCF in accordo con la BNCR. L’*harvesting* è una tecnica informatica per la raccolta dei dati tramite un apposito software, il *crawler*. In questo caso, l’*harvesting* viene fatto su repository che espongono i propri dati tramite OAI-PMH.¹² Tutte le informazioni tecniche per il deposito si trovano sul sito <http://www.depositolegale.it/>¹³ mentre l’elenco delle Università che aderiscono al servizio è consultabile su GitHub.¹⁴

L’*harvesting* viene effettuato con cadenza regolare, almeno una volta all’anno; fino al 2015 le ricevute dell’*harvesting*, in formato XML, venivano inviate tramite email ai competenti uffici d’ateneo ma per la difficoltà di gestire un indirizzario in continuo aggiornamento, l’invio delle ricevute è stato interrotto. È possibile verificare se una tesi è stata depositata o meno sul sito <https://index.depositolegale.it/>: la ricerca può essere effettuata a partire dall’indirizzo URL di una specifica tesi o sull’intero dominio di un’Università, ad es. *.unipd.it, ottenendo in quest’ultimo caso tutti gli indirizzi URL archiviati. Il report è scaricabile in TXT o in JSON, per facilitare l’elaborazione automatica, ma non risulta sufficientemente usabile per gli uffici amministrativi degli atenei. Per questo motivo, nel 2019 la BNCF metterà a disposizione un’API che indicizza e facilita il recupero, la visualizzazione e il download di un elenco di metadati minimali (autore e titolo, ateneo, data di pubblicazione, data dell’*harvesting*...) delle tesi di dottorato depositate. Da *Index* è possibile ricercare anche le URL delle tesi pervenute su CD e DVD-ROM poiché il repository locale della BNCF viene sottoposto ad *harvesting* come gli altri repository istituzionali: il suo dominio è <http://tesi.depositolegale.it>.

Nel complesso, dopo la sessione di luglio 2017,¹⁵ sono state raccolte oltre 76.000 tesi di dottorato. Le copie d’archivio sono accessibili da <https://memoria.depositolegale.it/>: la ricerca può essere fatta unicamente per URL della tesi e la visualizzazione è possibile solo dalla rete interna della BNCF. Questa limitazione, in analogia a quanto avviene per le tesi di dottorato cartacee, serve a restringere

¹⁰ <https://www.crui.it/linee-guida-per-gli-archivi-istituzionali.html>.

¹¹ <https://www.crui.it/linee-guida-per-la-creazione-e-la-gestione-di-metadati-nei-repository-istituzionali.html>.

¹² <https://www.openarchives.org/OAI/openarchivesprotocol.html>.

¹³ Il sito è in corso di aggiornamento ma la sezione relativa alle modalità di deposito delle tesi è sostanzialmente valida.

¹⁴ <https://github.com/depositolegale/sources>.

¹⁵ I dati del 2018 non sono ancora disponibili al momento della stesura del presente contributo. Il numero delle tesi raccolte, suddiviso per anno di cattura e ateneo, sono disponibili su GitHub: <https://github.com/depositolegale>.

le possibilità che la tesi venga illegalmente riprodotta, ma verrà probabilmente meno quando sarà in produzione il browser virtuale¹⁶ di Magazzini Digitali.

Un appunto in merito ai record presenti nei repository che subiscono variazioni nel tempo:¹⁷ il comportamento standard dei repository, come EPrints o DSpace, è di ripubblicare il record OAI-PMH a ogni modifica effettuata, variandone la data di modifica. L'*harvesting* viene effettuato in maniera incrementale, ciò vuol dire che, in ciascuna sessione, viene catturato tutto ciò che è nuovo o diverso rispetto a quanto già presente in archivio, senza sovrascritture o controlli sulle modifiche. Se quindi il record di una tesi è stato modificato, sarà possibile vedere in *Memoria* tutte le versioni catturate con la data di cattura.

I file WARC¹⁸ risultanti dall'*harvesting* saranno conservati in Magazzini Digitali.

Per completezza è anche necessario ricordare che la CRUI e Magazzini Digitali collaborano da molti anni per portare avanti il progetto NBN Italia (NBN:IT),¹⁹ grazie al quale alle risorse che entrano in Magazzini Digitali, comprese le tesi di dottorato, può essere attribuito su richiesta un identificatore persistente della risorsa, l'*NBN* (National Bibliography Number) appunto. Tramite il *resolver* messo a disposizione da depositolegale.it è possibile accedere alla copia della pubblicazione conservata in *Memoria*: <http://nbn.depositolegale.it/>.

Le policy d'accesso e i casi di embargo

Nonostante il D.P.R. n. 382/1980, fin da subito, abbia chiarito che le tesi di dottorato sono documenti di natura pubblica soggetti ad obbligo di deposito presso le due Biblioteche Nazionali Centrali che le rendono disponibili per almeno 30 anni,²⁰ le Università hanno adottato prassi talvolta difformi in materia di proprietà intellettuale e consultabilità delle tesi di dottorato.²¹ Solo l'adesione degli atenei

¹⁶ I browser virtuali (emulati) all'interno del browser permettono la normale navigazione delle risorse ma non consentono le operazioni di salvataggio e copia/incolla.

¹⁷ In linea di principio il full-text della tesi non dovrebbe mai essere modificato perché è sulla base della versione del testo archiviata nel repository che viene rilasciato il titolo di Dottore di ricerca. Viceversa, alcuni metadati possono essere modificati, soprattutto quando vengono riscontrati errori o lacune.

¹⁸ https://en.wikipedia.org/wiki/Web_Archive.

¹⁹ Per approfondimenti: <http://www.depositolegale.it/national-bibliography-number/> e Bellini (2012).

²⁰ Le tesi in formato cartaceo non sono mai state oggetto di scarto anche trascorsi 30 anni dalla loro pubblicazione, ma non vi è l'obbligo effettivo di conservazione dopo tale termine. Viceversa, le tesi in formato digitale per le regole stesse della long term preservation vengono conservate e rese disponibili a tempo indeterminato.

²¹ Sui repository istituzionali e le politiche Open Access è stato scritto molto e ai fini del presente contributo si ritiene non sia il caso di soffermarsi ulteriormente sull'argomento. Alcuni suggerimenti di lettura sono Guerrini (2010) e Solodovnik (2015). Un'ottima sintesi delle problematiche si trova anche nell'introduzione delle già citate "Linee guida della CRUI per il deposito delle tesi di dottorato negli archivi aperti" e nel documento "Tesi di dottorato e diritto d'autore" sempre a cura del Gruppo OA della CRUI: <https://www.crui.it/images/biblioteche/AddendaLineeGuidaTesi.pdf>.

L'elenco delle università italiane che hanno adottato una policy in materia di Accesso Aperto è consultabile sulla pagina Wiki di OA Italia: http://wikimedia.sp.unipi.it/index.php/OA_Italia/Regolamenti_e_Policy_sull%27Open_Access.

Per una panoramica più completa è possibile consultare il ROARMAP - Registry of Open Access Repository Mandates and Policies: <http://roarmap.eprints.org/>.

italiani alla Dichiarazione di Berlino²² in materia di Open Access ha sancito definitivamente il principio per cui i risultati della ricerca finanziata con fondi pubblici sono pubblici: il dottorando può conseguire il titolo solo se accetta che l'elaborato finale sia liberamente consultabile, e ciò si concretizza nell'obbligo di auto-archiviazione in accesso aperto della tesi nel repository istituzionale. Esistono comunque casi in cui è previsto un embargo temporaneo,²³ che deve essere rispettato anche per la copia di deposito: alcune Università non consentono la cattura delle tesi sotto embargo, altre la consentono ma in Magazzini digitali viene apposto un blocco²⁴ per cui i metadati della tesi sono ricercabili in *Index*, ma il full text non è visibile in *Memoria* fino al termine dell'embargo.

Strategie per la ricerca delle tesi di dottorato nei cataloghi della BNCF

Le tesi di dottorato pervenute in formato cartaceo in BNCF e in BNCR sono ricercabili negli OPAC delle due Biblioteche Nazionali Centrali e nell'OPAC SBN. Inoltre le tesi di dottorato pervenute in formato cartaceo costituiscono un'apposita serie della BNI - Bibliografia Nazionale Italiana,²⁵ i cui fascicoli a partire dalle annate 2012 sono liberamente consultabili online.²⁶ La catalogazione viene fatta in base alle regole SBN e alle norme per la redazione dei record BNI, e sono oggetto di catalogazione semantica; laddove possibile, nel record descrittivo si riporta il codice SSD (Settore Scientifico-Disciplinare) MIUR. Le tesi di dottorato in formato cartaceo sono consultabili unicamente presso le due Biblioteche Nazionali di Firenze e Roma. I record descrittivi delle tesi in formato elettronico, pervenute su CD o oggetto di *harvesting* da repository, non sono ancora disponibili nell'OPAC della BNCF²⁷ perché, in assenza del browser virtuale per la visualizzazione controllata dei file in *Memoria*, si è ritenuto più opportuno non esporre

²² Gli atenei italiani aderirono alla Dichiarazione di Berlino con la Dichiarazione di Messina del 2004: molte informazioni sono reperibili nel sito del decennale della sottoscrizione: <http://decennale.unime.it>.

²³ All'articolo 9 del già citato documento "Tesi di dottorato e diritto d'autore" a cura del Gruppo OA della CRUI si legge: "L'embargo è il periodo di tempo durante il quale la tesi archiviata in un deposito istituzionale risulta secretata ed accessibile solo per la parte dei metadati (il periodo di tempo può variare fra i sei e i dodici mesi oppure, in casi eccezionali, per un periodo massimo di tre anni). Il dottorando ha il diritto di chiedere un periodo di embargo, purché la richiesta sia debitamente motivata e giustificata. L'embargo si applica solo quando è opportunamente motivato".

²⁴ Attualmente il blocco viene configurato manualmente e solo sulle tesi per cui viene fatta esplicita richiesta dalle Università: in assenza di questa esplicita richiesta, la tesi catturata è potenzialmente consultabile perché i metadati con la data di fine dell'embargo non sono quasi mai espressi in modo uniforme e non è possibile applicare automatismi. A breve, dovrebbe essere disponibile un'implementazione del software pywb (Python Wayback), che sta alla base di memoria.depositolegale.it, e che consentirà di gestire in maniera più granulare whitelists/blacklists di URL ad accesso ristretto in repository ad accesso aperto. La modifica è stata commissionata dalla British Library: <http://blogs.bl.uk/webarchive/2018/02/a-new-playback-tool-for-the-uk-web-archive.html>.

Il codice di pywb è su GitHub: https://github.com/ukwa/ukwa-pywb/blob/master/docs/access_controls.md.

²⁵ In BNCF si sta valutando se e come dare notizia nella Bibliografia Nazionale anche delle tesi di dottorato pervenute in formato elettronico, attualmente non censite.

²⁶ Le diverse serie della BNI, a partire dall'annata 2012, sono consultabili sul sito: <http://bni.bncf.firenze.sbn.it/bniweb/menu.jsp>. Dalle annate 2015, i fascicoli sono scaricabili oltre che in formato .pdf anche in Unimarc e in XML.

²⁷ Nel 2013 fu eseguito un import nell'OPAC della BNCF, di cui si trovano ancora alcuni esempi, dei record Unimarc ottenuti a partire dal tracciato Dublin Core delle tesi elettroniche ma il progetto venne interrotto per la mancanza del browser virtuale.

le notizie. Al momento della messa in produzione del browser virtuale è previsto un import nell'OPAC di BNCf dei record descrittivi, che riporteranno i metadati esposti nei repository.²⁸ Attualmente gli strumenti per la ricerca delle tesi in formato elettronico sono quelli indicati nel paragrafo "L'*harvesting* delle tesi di dottorato". Per quanto riguarda la condivisione delle notizie con l'Indice SBN, doveroso o fortemente auspicabile, si dovrà procedere con cautela dal momento che per poter effettuare gli import è necessario un controllo preventivo sui record d'autorità, per evitare duplicazioni. Da questo punto di vista, le tesi di dottorato in formato elettronico costituiscono un banco di prova rispetto alla gestione complessiva del *born digital* in SBN: l'attuale architettura, non solo per i richiamati problemi relativi agli *authority file*, non è adatta a ospitare in maniera agile questi record "non SBN nativi" e i relativi oggetti digitali, seppure sia ormai indubbio che questi documenti siano una parte anche quantitativamente molto significativa del panorama informativo, destinata a crescere esponenzialmente.

Prospettive

L'Italia è probabilmente l'unico Paese in cui la pubblicità delle tesi di dottorato sia determinata *ex lege* e legata all'obbligo di deposito presso le Biblioteche Nazionali Centrali, che, oltre a renderle consultabili, ne curano la conservazione. Nel resto del mondo è d'obbligo unicamente il deposito della tesi da parte del dottorando presso la propria Università (come d'altronde in Italia), e quasi ovunque la pubblicità delle tesi deriva dall'assimilazione agli altri prodotti della ricerca scientifica. Le biblioteche si limitano, per così dire, a partecipare a progetti collaborativi per la creazione di sistemi di discovery oppure di veri e propri repository di repository con lo scopo di valorizzare e rendere fruibili queste risorse, in primo luogo per i ricercatori; progetti in cui spesso il supporto di aziende o consorzi privati è fondamentale. A titolo esemplificativo²⁹ si possono citare: *theses.fr*,³⁰ il discovery nazionale delle tesi di dottorato francesi mantenuto da ABES - Agence bibliographique de l'enseignement supérieur; *ETbOS - eThesis Online Service*³¹ della British Library; il discovery *Trove*³² della National Library of Australia; o ancora l'indiano *Shodhganga*,³³ repository e discovery mantenuto dall'INFLIBNET Centre - Information and Library Network Centre e l' *Open Dissertations database*³⁴ di EBSCO; infine, forse il caso più noto della Library of Congress che per i suoi utenti prevede l'accesso al database *Dissertations and Theses Global*³⁵ implementato e mantenuto da ProQuest. Anche per la ricerca delle tesi italiane esistono sistemi altri rispetto a Magazzini Digitali, più legate al mondo Open

²⁸ A questo proposito è interessante anche solo accennare al fatto che le tesi di dottorato, che tendenzialmente hanno un paratesto uniforme e un testo ricco di terminologia scientifica, potrebbero essere usate per il training dei sistemi di indicizzazione automatica.

²⁹ Una lista ampia ed aggiornata di iniziative nazionali ed internazionali per l'accesso alle tesi si trova sul sito della Library of Congress: <https://www.loc.gov/rr/main/alcove9/education/theses.html>.

³⁰ <http://www.theses.fr/>.

³¹ <https://ethos.bl.uk>.

³² <https://trove.nla.gov.au/>.

³³ <http://shodhganga.inflibnet.ac.in/>.

³⁴ <https://biblioboard.com/opendissertations/>.

³⁵ <https://www.proquest.com/libraries/academic/dissertations-theses/>.

Access: il più importante certamente è il portale *PLEIAD*³⁶ frutto della collaborazione tra i due consorzi interuniversitari CASPUR e CILEA, poi confluiti entrambi in CINECA; inoltre i repository di molte Università italiane vengono raccolti in *DART-Europe*³⁷ mantenuto da biblioteche e consorzi di biblioteche di ricerca europei.

In Italia l'aspetto della conservazione sembra aver preso il sopravvento su quello dell'accesso, probabilmente anche a causa di una normativa che non tiene conto delle mutate modalità di diffusione delle informazioni, ma non solo. Il Web ed i suoi standard evolvono in continuazione e molto velocemente, e le procedure di *Web archiving* devono adeguarsi a questa evoluzione. Se sul fronte degli strumenti di cattura e *replay* delle risorse archiviate, Magazzini digitali è riuscito a rimanere aggiornato, pur con tutte le difficoltà del caso, sappiamo che oggi esisterebbero tecnologie migliori per ciò che riguarda la sincronizzazione dei metadati (OAI-ORE³⁸ o Signposting³⁹) o la loro modellazione (Triple RDF e, eventualmente, Linked Data). Purtroppo, data l'eterogeneità dei repository e le difficoltà tecniche in cui operano le università, l'*harvesting* delle tesi resta legato all'uso di OAI-PMH, una soluzione priva di costi di implementazione e disponibile in tutti i software di repository digitali.

In conclusione, il deposito in Magazzini Digitali e l'attribuzione di un NBN alle tesi di dottorato garantiscono indubbiamente la conservazione a lungo termine, ma in questi anni non si sono potute investire le stesse risorse in sistemi di ricerca e accesso più immediati: in BNCf si sta lavorando per rendere anche questi più efficienti, facendo rientrare il deposito delle tesi di dottorato all'interno di una reingegnerizzazione complessiva di Magazzini Digitali.

References

- Allegrezza, Stefano. 2015. "L'affidabilità delle memorie ottiche per la conservazione a lungo termine dei documenti digitali." *JLIS.it* Vol. 6, n. 2 (2015). Doi: [jlis.it-11146](https://doi.org/10.4403/jlis.it-11146).
- Bellini, Emanuele *et al.* 2012. "The National Bibliography Number Italia (NBN:IT) Project. A persistent identifier supporting national legal deposit for digital resources." *JLIS.it* Vol. 3, n. 1 (2012). Doi: [jlis.it-4789](https://doi.org/10.4403/jlis.it-4789).
- Bergamin, Giovanni e Messina, Maurizio. 2010. "Magazzini digitali: dal prototipo al servizio." *Digitalia* Vol. 2 (2010). URL: <http://digitalia.sbn.it/article/view/246>.
- Galimberti, Paola. 2013. "Lunga vita alle tesi (di dottorato)." *Newsletter Roars Review* 13 febbraio 2013. URL: <https://www.roars.it/online/lunga-vita-alle-tesi-di-dottorato/>.
- Guerrini, Mauro. 2010. *Gli Archivi istituzionali: Open Access, valutazione della ricerca e diritto d'autore*. Milano: Editrice Bibliografica.

³⁶ <http://find.openarchives.it/>.

³⁷ <http://www.dart-europe.eu>.

³⁸ <https://www.openarchives.org/ore/>.

³⁹ <http://signposting.org/>.

Solodovnik, Iryna. 2015. *Repository Istituzionali Open Access e strategie Linked Open Data: Per una migliore comunicazione dei prodotti della ricerca scientifica*. Firenze: Firenze University Press. Accessibile anche all'URL: http://www.fupress.com/archivio/pdf/3091_9271.pdf.

Shirazi, Roxanne. 2018. "The doctoral dissertation and scholarly communication. Adapting to changing publication practices among graduate students." *College & Research Libraries News* Vol. 79 n. 1 (2018). URL: <https://crln.acrl.org/index.php/crlnews/article/view/16864/18489>.

Vignocchi, Maria Laura, Bergamin, Giovanni e Messuti, Raffaele. 2010. "Tesi di dottorato: stato dell'arte, iniziative in corso, scenari possibili." *Bibliotime* anno XIII, n. 3 (novembre 2010). URL: <http://www.aib.it/aib/sezioni/emr/bibtime/num-xiii-3/vignocchi.htm>.

Vignocchi, Maria Laura e Lauriola, Roberta. 2011. "Il deposito legale delle tesi di dottorato di ricerca: l'esperienza dell'Università di Bologna." *Digitalia* Vol. 2 (2011). URL: <http://digitalia.sbn.it/article/view/482>.